Social Choice and Game Theory in Allocation Mechanisms

Lindeneg, Klavs

Publication date:
2000

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Lindeneg, K. (2000). *Social Choice and Game Theory in Allocation Mechanisms*. (e-publikation, rev. 2001).

Social Choice and Game Theory in Allocation Mechanisms

Klavs Lindeneg

June 18, 2001

Contents

1
Introduction

In the theory of implementation, see e.g. Corchón (1996), social choice theory and game theory are used intensively. These lecture notes survey very briefly some elements of social choice theory and game theory used in a course on Allocation Mechanisms at the University of Copenhagen, Institute of Economics, basically non-cooperative, non-Bayesian implementation of some socially optimal allocations in economic models.

Social choice theory and game theory are concerned with problems of collective decision-making: several agents have to decide on some issue of collective interest whereas their preferences about the issue and/or their endowments might differ. In the theory of implementation normative (prescriptive value judgements as represented in social choice mappings) as well as descriptive (strategic behavior as represented by game-theoretical equilibrium concepts) features of collective decision making are involved. The concept of implementation also proves to be central for describing the relationship between the normative and positive properties of, and the use of social choice theory and game theory in collective decision making: social choice theory answer the feasibility problem whether or not a particular (ethic) social choice mapping can be materialized, e.g. in an economy of selfish individual agents, while game theory tell us how a feasible social choice mapping can be implemented by decentralizing the decision power via a particular game form, according to some pattern of behavior (equilibrium concept).

Writings in social choice theory dates back to political philosophers in the late eighteenth century, e.g. Borda (1781) and Condorcet (1785). Later the axiomatic method was introduced in the theory by Arrow (1951) and May (1952). In the following we present some impossibility results from the latter period.

The break-through for game theory in economics was the book by von Neumann and Morgenstern (1944). We present some non-cooperative, non-Bayesian equilibrium concepts and some associated existence results.

The material in these notes is collected from the following textbooks and lecture notes: Moulin (1983), Keiding (1984), Keiding (1987), Gibbons (1992) and Mas-Colell *et al.* (1995). In chapter 2 on social choice theory some difficulties in making collective decisions are stated: (2.1) Arrow's impossibility theorem saying that it is impossible to treat a collective of people as a single agent having a total preorder of the feasible alternatives if the choice is not binary; (2.2) the Gibbard-Satterthwaite theorem saying that strategic behavior is an immanent part of collective decisions, if the choice is not binary, and if a single alternative must be chosen by a plurality of agents with different preferences; and (2.3) that reasonable collective choices are possible to achieve if only a subset containing more than one feasible alternative must be chosen. In chapter 3 on game theory we look a bit more deep into the strategic possibilities for the agents in collective decision making: (3.1) dominating strategies; (3.2) Nash-strategies; (3.3-3.4) some refinements of Nash-strategies. And in chapter 4 we characterize very briefly some allocation mechanisms.

2
Social Choice Theory

In social choice theory we look at what reasonable outcomes of collective decisions it will be possible to obtain. We assume a world where individuals are characterized solely by their preferences, everybody's preferences are publicly known (until section 2.2.1), and formalize the collective decision problem by a mapping that selects for any given profile of preferences a social welfare function (a collective preference ordering) or a subset of outcomes (a social choice correspondence or a social choice function if the subset is a singleton). Such a mapping summarizes a particular ethic of collective decisions, a compromising rule applicable to every possible configuration of individual preferences.

Throughout we adhere to the assumption that both the outcomes and the agents participating in collective decisions are finitely many. And the individual preferences are assumed to be linear orders, i.e. indifferences are ruled out.

Let us assume we have a set of rational individuals, $M = \{1, ..., m\}$, each of them having their own total linear preorder on the set of feasible alternatives, $A = \{1, ...p\}$, from which they collectively can choose. If the set of linear preorders over A is called $L(A)$, we have the following

Definition 1 *A total linear preorder for agent i,* $\succ_i \in L(A)$ *is antisymmetric, transitive and total, i.e.* $\forall i \in M : \succ_i$ *is*

(A1) Antisymmetric: $\forall a, b \in A$: *if* $a \succ_i b$ *and* $b \succ_i a$*, then* $a = b$*,*

(A2) Transitive: $\forall a, b, c \in A$: *if* $a \succ_i b$ *and* $b \succ_i c$ *then* $a \succ_i c$*,*

(A3) Total: $\forall a, b \in A : a \succ_i b$ *or* $b \succ_i a$*.*

- This means that, indifference are ruled out (A1), and (A2)-(A3) all individuals are rational (they have transitive preferences and all the information on the set of alternatives they need to make a rational choice).[1] An example of a linear preorder is the real numbers, $\mathbb{R} \in L(A)$.

Definition 2 *A preference profile,* $(\succ_1, ..., \succ_m)$*, is a mapping from* M *to* $L(A)^M$*, assigning to each individual a linear preorder* $\succ_i \in L(A)$*.*

A collective decision problem exist, if for some $(\succ_i, \succ_j) \subset L(A)^M : \succ_i \neq \succ_j$. Since A is finite, each $\succ_i \in L(A)$ can be represented by a utility function, $u_i : A \curvearrowright \mathbb{R}$, if and only if $a \succ_i b \Leftrightarrow u_i(a) > u_i(b)$ for all $\{a, b\} \subset A$. In many situations this way of describing preferences is more convenient, and therefore we will write profiles as $(\succ_1, ..., \succ_m)$ or $(u_1, ..., u_m)$ without further comment.

In the following we look at the feasibility of making mappings from preference profiles to the set of alternatives. In welfare economics this has been done variously - here we look at social welfare functions, social choice functions and social choice correspondences.

[1] Linear preorders are not used much in economics. But in our situation the set of alternatives is finite. Thus, linear orders can be considered as a good approximation to possibly more general preference relations. The argument for using linear orders here is that it is much more easy than using weak orders.

2.1 Social Welfare Functions

A social welfare function, swf, is a mapping from profiles of preferences, $L(A)^M$, to a social order of the alternatives in $L(A)$. A social order of alternatives is a mapping from $L(A)$ to A. Consequently, if we combine these two mappings, we have a mapping from preference profiles to the set of alternatives.

Definition 3 *A swf is a mapping* $w : L(A)^M \curvearrowright L(A)$*, where* $w(\succ_1, ..., \succ_m) = (\succ_{swf}) \in L(A)$ *is a procedure to aggregate preferences into a total linear preorder of* A*.*

An example is the utilitarian swf: $U_{swf}(a) = \sum_{i \in M} u_i(a)$ for all $a \in A$ (where each individual preference is represented by a cardinal utility function having the same units of utility for all agents).

In definition 3 a swf is a total linear preorder of the alternatives in A that satisfy a minimal condition of respect for individual preferences.

Is every social order a swf? No, e.g. pairwise majority voting is not (the Condorcet Paradox). Three agents (1,2,3) choose among three alternatives (a, b, c). The preferences of the agents are as follows

$$\begin{aligned} a &\succ {}_1 b \succ_1 c \\ c &\succ {}_2 a \succ_2 b \\ b &\succ {}_3 c \succ_3 a \end{aligned}$$

Then pairwise majority voting tells us, that a is socially preferred to b (by two votes to one), b is socially preferred to c (by two votes to one), and c is socially preferred to a (by two votes to one). But this cyclic pattern violates the transitivity requirement for a swf.

In this section we are looking for the possibilities of designing swf's satisfying some more conditions.

Definition 4 *A swf, w, is Paretian if, for any pair of alternatives, $\{a,b\} \subset A$, and any preference profile, $(\succ_1, ..., \succ_m) \in L(A)^M$, $w(a,b) = a$ whenever $\forall i \in M : a \succ_i b$.*

- A Paretian swf satisfies a minimal form of positive responsiveness - if everybody agree on a is strongly preferred to b, then the swf must choose a, $w(a,b) = a$.

Does a Paretian swf exist? Yes, an interesting example is the Borda Count. Suppose the number of alternatives in A equals p. Given an individual preference, $\succ_i$, we assign numbers to the p alternatives, e.g.[2] as follows: The most preferred alternative for every single agent is given p points, the second most preferred alternative is given $p-1$ points etc. Thus an alternative, a, is given a number by agent i equal to $c_i(a)$ - the better preferred the higher the number. Finally, for any profile of preferences, we determine the social ordering by adding up points to every alternative in A. By maximizing the score we get, if $\sum_{i \in M} c_i(a) > \sum_{i \in M} c_i(b)$ then $w(a,b) = a$. The Borda Count is Paretian, since if $a \succ_i b$ for all $i \in M$, then $c_i(a) > c_i(b)$ for all $i \in M$, and so $\sum_{i \in M} c_i(a) > \sum_{i \in M} c_i(b)$ - and then $w(a,b) = a$.

Is every swf Paretian? No, try to minimize instead of maximizing the Borda Count score.

Next we state a substantial restriction on swf's due to Arrow (1951), saying that the social preference of any two alternatives depends only on the individual preferences between the same two alternatives.

Definition 5 *A swf satisfies independence of irrelevant alternatives, if for all $\{a,b\} \subset A$ and any pair of preference profiles $(\succ_1, ..., \succ_m)$ and $(\succ_1^{'}, ..., \succ_m^{'})$ in $L(A)^M$ with the property that for every $i : a \succ_i b \Leftrightarrow a \succ_i^{'} b$ and $b \succ_i a \Leftrightarrow b \succ_i^{'} a$, we have $w(a,b) = a \Leftrightarrow w^{'}(a,b) = a$ and $w(a,b) = b \Leftrightarrow w'(a,b) = b$.*

[2] This can be done variously: in general elections every voter have one point to cast on one of all the candidates nominated for the parliament (an exeption is the election to the U.S. Senate in year 2000, where the winner in one state was a dead man, whose wife then took his seat); in the European Melody Contest every voter is given 67 points of which she must cast 12 on the best song, 10 on the second-best song, 9 on the third-best etc. But these - note the exeption - and all various ways to cast natural numbers on alternatives are examples of the type of voting procedure called the Borda count.

- Only the agents' preferences on a and b counts when a swf orders $\{a, b\} \subset A$. The justifications for this restriction are threefold. The first argument is strictly normative: in a social ranking between a and b, the presence or absence of alternatives other than a and b should not matter. Other alternatives are irrelevant to the issue in hand. The second is practically: the assumption facilitates the task of making social decisions because it helps to separate problems - the determination of a social ranking on a subset of alternatives does not need any information of individual preferences on the alternatives outside this subset. The third relates to incentives: independence of irrelevant alternatives is related to the feasibility of truthful revelation of individual preferences, see section 2.2.1 below.

In fact this restriction rules out some swf's, e.g. the Borda Count. The reason is, that the rank of an alternative in the Borda Count depends on the placement of every other alternative. If we have three alternatives and two agents with the preferences:

$$a \succ_1 b \succ_1 c$$
$$c \succ_2 a \succ_2 b$$

we have $\sum_{i=1,2} c_i(a) = 5$, and $\sum_{i=1,2} c_i(c) = 4$, and therefore $w(a, c) = a$. But for the preferences

$$\begin{aligned} a &\succ {}_1^{'} c \succ_1^{'} b \\ c &\succ {}_2^{'} b \succ_2^{'} a \end{aligned}$$

we have $\sum_{i=1,2} c_i(a) = 4$, and $\sum_{i=1,2} c_i(c) = 5$, and therefore $w(a, c) = c$. Yet the relative ordering of the two alternatives (a, c) has not changed for either of the two agents. It is the placement of the irrelevant alternative b that disorder the outcome of the Borda Count.

We commented on definition 3 of a swf, $w(\succ_1, ..., \succ_m)$, that a swf is a total linear preorder of the alternatives in A that satisfies some minimal condition of respect for individual preferences. Here we look at how minimal this respect can be.

Definition 6 *A dictator exists, if there* $\exists\, i \in M \mid \forall\, \{a,b\} \subset A$, *and* $\forall(\succ_1, ..., \succ_m) \in L(A)^M : a \succ_i b \Rightarrow w(a,b) = a$.

- If the swf always (in any pairwise comparison of alternatives in A) prefers the same alternative, as a specific individual in the society prefers, then this individual is called a dictator, and the swf is dictatorial. We then have

Theorem 1 *(Arrow (1951,1963). If* $|A| \geqq 3, (\succ_1, ..., \succ_m) \in L(A)^M$*, then every Paretian swf,* $w : L(A)^M \curvearrowright L(A)$*, that satisfied the independence of irrelevant alternatives is dictatorial (Arrow's Impossibility Theorem).*

Proof: We begin with defining a decisive subset of agents, $S \subset M$, given a swf, w :

(i) S is *decisive for* a *over* b, if whenever every agent in S prefers a over b, and every agent in $M \setminus S$ prefers b over a, then $w(a,b) = a$.

(ii) S is *decisive* if, for any pair $\{a,b\} \subset A$, S is decisive for a over b.

(iii) S is *completely decisive* for a over b, if whenever every agent in S prefers a over b, then $w(a,b) = a$.

The proof is done in a number of small steps. Steps 1-3 show that if a subset of agents is decisive for some pair of alternatives (i), then it is decisive for all pairs (ii). Steps 4-8 show that the smallest decisive set is formed by a single agent. And steps 9-10 show, that this agent is a dictator.

Step 1: *If for some* $\{a,b\} \subset A$*,* $S \subset M$ *is decisive for* a *over* b*, then, for any alternative* $c \in A$*,* $c \neq a$*,* S *is decisive for* a *over* c. If $c = b$ there is nothing to prove, so assume $c \neq b$. Consider a profile in $L(A)^M$ where

$$a \succ_i b \succ_i c \text{ for } \forall i \in S \subset M$$

$$b \succ_i c \succ_i a \text{ for } \forall i \in M \setminus S.$$

Then, because (i) S is decisive for a over b, we have $w(a,b) = a$. In addition, since $\forall i \in M : b \succ_i c$, we have by the Paretian assumption that $w(a,c) = a$. And by the independence of irrelevant alternatives, it follows that $w(a,c) = a$ whenever every agent in S prefers a to c and every agent in $M \setminus S$ prefers c to a. Result, S is decisive for a over c.

Step 2: If S is decisive over $\{a,b,c\} \subset A$, then S is decisive for $\{a,b,c,d\} \subset A$. By step 1, if S is decisive for $\{a,b\}$ it is decisive for a over c and for b over c. But then, applying step 1 again, this time over the pair $\{a,c\}$ and a new alternative $d \in A$, we conclude that S is decisive over $\{c,d\}$. Similarly, applying step 1 on $\{b,c\}$ and d we get the same conclusion.

Step 3: If for some $\{a,b\} \subset A$, $S \subset M$ is decisive for a over b, then S is decisive (ii).This follows directly from step 2, where c and d can be any alternative in $A \setminus \{a,b\}$. Result, S is decisive for any pair of alternatives in A.

Step 4: If $S \subset M$ and $T \subset M$ are decisive, then $S \cap T$ is decisive. Take any triple of distinct alternatives in A and consider a profile of preferences in $L(A)^M$ where

$$\begin{aligned} c &\succ {}_i b \succ_i a \text{ for } \forall i \in S \setminus (S \cap T) \\ a &\succ {}_i c \succ_i b \text{ for } \forall i \in S \cap T \\ b &\succ {}_i a \succ_i c \text{ for } \forall i \in T \setminus (S \cap T) \\ b &\succ {}_i c \succ_i a \text{ for } \forall i \in M \setminus (S \cup T) \end{aligned}$$

Then $w(b,c) = c$, because S is a decisive set, and $w(a,c) = a$, because T is a decisive set. Therefore, by the transitivity of w, we have $w(a,b) = a$. It follows by independence of irrelevant alternatives that $S \cap T$ is decisive for $\{a,b\}$, and so - by step 3 - that $S \cap T$ is a decisive set.

Step 5: For any $S \subset M$, either S or $M \setminus S \subset M$ is decisive. Again, take any triple of distinct alternatives in A, and consider a profile of preferences in $L(A)^M$ where

$$\begin{aligned} a \succ_i c \succ_i b &\quad \text{for } \forall i \in S \\ b \succ_i a \succ_i c &\quad \text{for } \forall i \in M \setminus S \end{aligned}$$

Then $w(a,b) = a$, if S is decisive (independence of irrelevant alternatives, and step 3 above), or $w(a,b) = b$. Because the

Paretian condition yields $w(a,c) = a$, the transitivity of w in this case gives $w(b,c) = b$. But then, using the independence of irrelevant alternatives, we conclude that $M \setminus S$ is decisive for $\{b,c\}$ (see the profile above), hence by step 3 decisive.

Step 6: If $S \subset M$ is decisive, and $S \subset T \subset M$, then T is decisive. By the Paretian condition the empty set of agents cannot be decisive. Therefore $M \setminus T$ cannot be decisive, because otherwise, by step 4, $S \cap (M \setminus T) = \emptyset$ would be decisive. Hence, by step 5, T is decisive.

Step 7: If $S \subset M$ includes more than one agent and is decisive, then there is a decisive strict subset $S' \subset S$. Take any $i \in S$. If $S \setminus \{i\}$ is decisive, then we are done. If $S \setminus \{i\}$ is not decisive, then, by step 5, $M \setminus S \cup \{i\}$ is decisive. When S and $M \setminus S \cup \{i\}$ are both decisive, then, by step 4, $\{i\}$ is decisive. And, by assumption, $\{i\}$ is a strict subset of S.

Step 8: There is an $i \in M$ such that $S = \{i\}$ is decisive. Because the set M is finite, and, by the Paretian condition, decisive, this follows by iterating step 7.

Step 9: If $S \subset M$ is decisive then, for any $\{a,b\} \subset A$, S is completely decisive (iii). We want to prove that, for any $T \subset M \setminus S$, $w(a,b) = a$ whenever $\forall i \in S$ prefer a to b, every agent in T regards a better than b, and every other agent in M prefers b to a. Because of step 6, $S \smile T$ is decisive if S is decisive, and then we are done.

Step 10: If, for some $i \in M$, $S = \{i\}$ is decisive, then i is a dictator. If i is decisive, by step 9, i is completely decisive for any a over any b. This is, for $\forall \{a,b\} \subset A$, if $a \succ_i b$, then $w(a,b) = a$, and this is the definition of a dictator (definition 6 above).

If you combine steps 8 and 10 the proof of theorem 1 is completed. $\square$

Arrow's impossibility theorem[3] shows, that we should not expect a collective of individuals to behave like individuals. Even if you accept utility functions in the description of individual agents, you should not do so in the description of collectives of individuals, unless you will accept a dictator. Of course, there is exceptions from this conclusion, e.g. if we restrict the domain for individual preferences, see e.g. Mas-Colell *et al.* (1995) Chapter 21.D.

Instead of focusing on restricted domains we will in what follows look at other types of mappings from preference profiles to the set of alternatives, namely social choice functions and social choice correspondences.

2.2 Social Choice Functions

Instead of aggregating profiles of individual preferences into a rational social preference order, which then is used to make decisions in the society, in this and the next section we focus directly on social decisions. Thus we go all the way from preference profiles to alternatives in one step. In this section we use social choice functions, scf's, to pick up subsets of outcomes consisting of a singleton, $a \in A$.

Definition 7 *A social choice function, scf, is a mapping,* $f : L(A)^M \curvearrowright A$*, which assigns a chosen element* $f(\succ_1, ..., \succ_m) \in A$ *to any profile of individual preferences in* $L(A)^M$.

If A is finite, every swf on $L(A)^M$ induces a natural scf by associating with each $(\succ_1, ..., \succ_m) \in L(A)^M$ a most preferred element in A. In the Borda Count in section 2.1 above, the choice for every preference profile in $L(A)^M$ is an alternative, $a \in A$, that $\max\limits_{a \in A} \sum_{i \in M} c_i(a)$. Hence, in studying scf's, it is obvious to

[3] Arrow (1951, 1963) stated the theorem as a possibility theorem - indeed it is possible to construct a dictatorian swf - but in the 1970'ties, when it became more and more unacceptable to rule a society by a dictator, economists turn this possibility theorem into an impossiblity theorem: there exist no acceptable swf.

look for a result parallel to Arrow's impossibility theorem in section 2.1 above.

Recall that in Arrow's theorem we had two conditions: the swf had to be Paretian and independent of irrelevant alternatives. When we deal with scf's, again we require two conditions: the scf must be weakly Paretian and monotone.

Definition 8 *A scf, $f : L(A)^M \curvearrowright A$, is weakly Paretian if for any profile, $(\succ_1, ..., \succ_m) \in L(A)^M$, the choice, $f(\succ_1, ..., \succ_m) \in A$, is a weak Pareto optimum, if $\forall i \in M \mid \exists \{a, b\} \subset A : a \succ_i b \Rightarrow b \neq f(\succ_1, ..., \succ_m)$.*

Compared to the definition of a Paretian swf (definition 4 above), in definition 8 the scf is "weakly" Paretian, because the society does not have to choose a, even if everybody agrees that a is preferred to b, but on the other hand the society must not choose b in this case.

Definition 9 *A scf, $f : L(A)^M \curvearrowright A$, is monotone, if for any two profiles, $(\succ_1, ..., \succ_m)$ and $(\succ_1^{'}, ..., \succ_m^{'})$ in $L(A)^M$, with the property that if the chosen alternative, $a = f(\succ_1, ..., \succ_m)$, maintains its position in $(\succ_1^{'}, ..., \succ_m^{'})$, we have that $f(\succ_1^{'}, ..., \succ_m^{'}) = a$.*

- If no alternative can be dropped from being chosen, unless for some agents, its desirability deteriorates, then a scf is monotone.

As a corollary to Arrow's impossibility theorem, we have the following:

Corollary 2 *If $|A| \geq 3$ then every weakly Paretian and monotone scf is dictatorial.*

Proof: see Mas-Colell *et al.* (1995), proof of proposition 21.E.1.

2.2.1 Manipulation of Social Choice Functions

We now take a new look at the social choice procedure. Agents announce preferences, but there is actually no compelling reason to believe, or any way to enforce that the preferences announced are the agents true preferences. As we shall see later on in this section, there is every reason to believe that agents will not stick to their true preferences.

By a seemingly innocuous strengthening of monotonicity we get strong monotonicity.

Definition 10 *Given A and M, a scf, $f : L(A)^M \curvearrowright A$, is strongly monotone, if $(\succ_1^{'}, ..., \succ_m^{'})$ is obtained from $(\succ_1, ..., \succ_m)$ by an improvement of $a \in A$, then $f(\succ_1^{'}, ..., \succ_m^{'}) \in f(\succ_1, ..., \succ_m)$ $\cup \{a\}$.*

- That "$(\succ_1^{'}, ..., \succ_m^{'})$ is obtained from $(\succ_1, ..., \succ_m)$ by an improvement of $a \in A$" means that every alternative not involving a is ranked by everyone in $(\succ_1^{'}, ..., \succ_m^{'})$ as in $(\succ_1, ..., \succ_m)$; also, a is ranked at least as high by everyone in $(\succ_1^{'}, ..., \succ_m^{'})$ as in $(\succ_1, ..., \succ_m)$. Then the strong monotonicity assumption tells us that, in this case, the only alternative not chosen by $(\succ_1, ..., \succ_m)$ that can be chosen by $(\succ_1^{'}, ..., \succ_m^{'})$ must be a.

At the surface this assumption seems uncontrovertial: if everybody in society changes his/her view on a in a positive manner - or just somebody does, and the preferences of the rest remain unchanged - then it is quite reasonable that society should reflect this movement of alternative a in its choice. However, the implications of this assumption are far-reaching, see theorem 3 below. But we need another assumption to get the theorem:

Definition 11 *Given A and M, a scf, $f : L(A)^M \curvearrowright A$, satisfies citizens' sovereignty if, for all $a \in A$, there is a profile of preferences, $(\succ_1, ..., \succ_m) \in L(A)^M$ with $f(\succ_1, ..., \succ_m) = a$.*

- Every alternative in A can be selected by the scf.

Then we have the following result:

Theorem 3 *(Muller and Satterthwaite (1977)). Let $|A| \geqq 3$, and $f : L(A)^M \curvearrowright A$ be a strong monotone scf satisfying citizens' sovereignty. Then f is dictatorial, i.e. there is an $i \in M : f(\succ_1, ..., \succ_m) = \max_{a \in A} u_i(a)$.*

Proof: For any $\{a, b\} \subset A$ define a map $G(a, b) : L(A)^M \curvearrowright L(A)^M$, taking each profile of preferences $(\succ_1, ..., \succ_m)$ to the profile $(\succ_1^{'}, ..., \succ_m^{'}) = G(a,b)(\succ_1, ..., \succ_m)$ with $\forall i \in M : x_1 \succ_i x_2$ if $x_1 \in \{a,b\}$, $x_2 \in A \setminus \{a,b\}$, and $x_1 \prec_i x_2 \Leftrightarrow x_1 \prec_i^{'} x_2$ otherwise. Thus $G(a,b)$ puts a and b in the top of the profile, while keeping the ranking of every other pair of alternatives unchanged.

Now we need what is called a generalized scf, F - e.g. F can be a rule assigning to each profile $(\succ_1, ..., \succ_m)$ a lottery $F(\succ_1, ..., \succ_m)$ on the profiles, with $F(\succ_1, ..., \succ_m)(P_{a>b})$ as the probability of choosing a linear order where a is preferred to b. We have the following:

Definition 12 *By $P_{a>b}$ we denote the set of all linear preorders, $P \in \Pi$, such that for $\{a,b\} \subset A : aPb$. A generalized scf is then a map, $F : L(A)^M \curvearrowright [0;1]$, taking profiles into functions from Π to $[0;1]$,*

$$F(\succ_1, ..., \succ_m)(P_{a>b}) = \begin{cases} 1 & \textit{if } f(G(a,b)(\succ_1, ..., \succ_m) = a \\ 0 & \textit{otherwise} \end{cases}$$

Then F is binary:

Definition 13 *Let F be a generalized scf. F is binary if for all $\{a,b\} \subset A$, $(\succ_1, ..., \succ_m)$ and $(\succ_1^{'}, ..., \succ_m^{'}) \in L(A)^M : (\succ_1, ..., \succ_m)^{-1}(P_{a>b}) = (\succ_1^{'}, ..., \succ_m^{'})^{-1}(P_{a>b}) \Rightarrow F(\succ_1, ..., \succ_m)(P_{a>b}) = F(\succ_1^{'}, ..., \succ_m^{'})^{-1}(P_{a>b})$.*

Now it is easy to show, that $f(G(a,b)(\succ_1, ..., \succ_m)) = \{a,b\}$: by citizens' sovereignty there is a $(\succ_1, ..., \succ_m)^a$ with $f(\succ_1, ..., \succ_m)^a = a$, and since $G(a,b)(\succ_1, ..., \succ_m)^a$ is obtained from $(\succ_1, ..., \succ_m)^a$ by an improvement of a and b, we must have $f(G(a,b)(\succ_1, ..., \succ_m) \in \{a,b\}$. If then we change the position of only a and b we can get a profile $(\succ_1^{'}, ..., \succ_m^{'})^{-1}(P_{a>b}) = G(a,b)(\succ_1, ..., \succ_m)^{-1}(P_{a>b})$, and from $f(\succ_1^{'}, ..., \succ_m^{'}) \in \{a,b\}$ we conclude that a binary $f(G(a,b)(\succ_1, ..., \succ_m)) \in \{a,b\}$.

Furthermore, let F satisfy conditional monotonicity:

Definition 14 *Let F be a binary generalized scf. F is conditionally monotone if for all $\{a,b,c\} \subset A$, $(\succ_1, ..., \succ_m)^{-1}(P_{a>b}) = m$ then $F(\succ_1, ..., \succ_m)(P_{b>c}) \leq F(\succ_1, ..., \succ_m)(P_{a>c})$.*

A strong monotone generalized scf is conditionally monotone (saying that conditionally monotonicity is a weaker condition). If $(\succ_1, ..., \succ_m)^{-1}(P_{a>b}) = m$, then $F(\succ_1, ..., \succ_m)(P_{a>b}) = 1$. And suppose further that $F(\succ_1, ..., \succ_m)(P_{b>c}) = 1$ meaning that $f(G(a,b)(\succ_1, ..., \succ_m)) = b$. Define another profile, $(\succ_1^{'}, ..., \succ_m^{'})$, by an improvement of a in $G(a,b)(\succ_1, ..., \succ_m)$ such that a is amongst the top three alternatives, and $\{a, b, c\}$ are ranked as in $(\succ_1, ..., \succ_m)$. Then $f(\succ_1^{'}, ..., \succ_m^{'}) \in \{a, b\}$ by strong monotonicity. If in a new profile, $(\succ_1^{''}, ..., \succ_m^{''})$, c is moved down from the top three places, then a will be chosen (if the profile is strong monotone and satisfied citizens' sovereignty). But this imply that $f(\succ_1^{'}, ..., \succ_m^{'}) = a$ since $(\succ_1^{'}, ..., \succ_m^{'})$ is obtained from $(\succ_1^{''}, ..., \succ_m^{''})$ by an improvement of c. And it is clear that moving b down in the profile results in $f(G(a,b)(\succ_1, ..., \succ_m)) = a$, and $F(\succ_1, ..., \succ_m)(P_{a>c}) = 1$.

We leave it to the reader to check steps 4-10 in theorem 1 above: the smallest decisive set is formed by one agent, and that this agent is a dictator. Then we are done. □

According to theorem 3 there are severe limitations to the possibilities of constructing scf's. But instead of dealing with them here, we state another theorem which is far-reaching in its simplicity - we get it as a corollary to theorem 3 above. Before the theorem is stated we need another definition.

Definition 15 *A scf, $f : L(A)^M \curvearrowright A$, is strategy-proof, if for each agent, $\forall\ i \in M$, there exist a mapping, $\succ_i : L(A) \curvearrowright A$, such that for every agent we have: $\forall \succ_i \in L(A), \forall \succ_{)i(} \in L(A)^{M \setminus \{i\}}, \forall \succ_i^{'} \in L(A) : f(\succ_i, \succ_{)i(}) \succ_i f(\succ_i^{'}, \succ_{)i(})$, where $\succ_i$ is the agent's sincere preferences.*

- In a strategy-proof scf "telling the truth" is always a dominating strategy for every agent, who can simply ignore the behavior of the other agents when he chooses his strategy.

Theorem 4 *(Gibbard (1973) and Satterthwaite (1975)). Let $|A| \geqq 3$, and $f : L(A)^M \curvearrowright A$ be a scf satisfying (definition 11 above) citizens' sovereignty. Then the scf is strategy-proof if and only if it is dictatorial.*

Instead of scf's this theorem can be shown to hold for game forms, see Moulin (1983, chapter 4). It tells, that strategic behavior is part of collective decisions, if the result must be a singleton, $a \in A$, if the decision is not binary, and if the choice solely depends on the preference profile of the decision-makers.

Proof: The "if" statement is clear: if a dictator decides in the name of the whole society then it will be a dominating strategy for the dictator to tell the truth, $\succ_i$ in definition 15. Conversely, we will prove that a strategy-proof scf is strongly monotonic. And then the Muller-Satterthwaite theorem above complete the proof.

Let f be a scf which is not strongly monotonic. Then there exist an $a \in A$, and profiles $(\succ_1, ..., \succ_m)$ and $(\succ_1^{'}, ..., \succ_m^{'}) \in L(A)^M$ such that $(\succ_1^{'}, ..., \succ_m^{'})$ is obtained from $(\succ_1, ..., \succ_m)$ by improvement of a, whereas $f(\succ_1^{'}, ..., \succ_m^{'}) \nsubseteq \{a, f(\succ_1, ..., \succ_m)\}$, e.g. "outcome a jumps above outcome b in agent $i's$ ordering when we go from $(\succ_1, ..., \succ_m)$ to $(\succ_1^{'}, ..., \succ_m^{'})$". If $f(\succ_1, ..., \succ_m) = c$, $f(\succ_1^{'}, ..., \succ_m^{'}) = d$, and $c \neq d$, we distinguish two cases:

(i) $\{c, d\} \neq \{a, b\}$. Then the relative ordering of c and d in $(\succ_1, ..., \succ_m)$ and $(\succ_1^{'}, ..., \succ_m^{'})$ is the same, so one of the two following inequalities holds: $d \succ_i c$ (i.e. $f(\succ_i^{'}, \succ_{)i(})_{i \in M} \succ_i f(\succ_i)_{i \in M}$) or $c \succ_i^{'} d$ (i.e. $f(\succ_i)_{i \in M} \succ_i^{'} f(\succ_i^{'}, \succ_{)i(})_{i \in M}$) - in each case f violates strategy-proofness.

(ii) $\{c, d\} = \{a, b\}$. Since $d \neq a$ we have $a = c$ and $b = d$. Then $f(\succ_i^{'}, \succ_{)i(})_{i \in M} = b \succ_i f(\succ_1, ..., \succ_m) = a$, and again strategy-proofness of f is violated. $\square$

Theorem 4 shows that only binary choices can actually be made via strategy-proof collective decisions methods, if the result must be a singleton.

Several ways exist to escape the Gibbard-Satterthwaite result:

(a) Restrict the domain of feasible preference profiles.

(b) Weaken the equilibrium concept and the requirement that a dominant strategy equilibrium exists for all profiles.

(c) Accepting results of collective decisions which are not singletons, but subsets of alternatives in A containing more than one alternative.

In the following we will look at possibilities in the line of (b) in section 3 below, and (c) in section 2.3.

2.3 Social Choice Correspondences

If we accept results of a collective decision which are subsets of alternatives in A containing more than one single alternative, we seemingly have chosen not to solve all collective decision problems. But in economic theory, and particular in welfare economics, it is more the rule than an exception to look at correspondences, e.g. the Pareto-correspondence, the envy-free correspondence etc.

One reason for the relevance of social choice correspondences in social choice theory is, that we have restricted ourselves to use information only of one single characteristic of every agent, the agents preferences; and then it is difficult (on the domain of preference profiles) to choose one single alternative in A. What specific alternative an agent will choose depends in welfare economics, besides on his preferences, also on the agent's wealth (capital and income). Therefore it does not matter if we cannot find a singleton in A, if our goal is to construct "the rules of a game" which will be played over and over again by different players in different (economic) situations (capital, income, consumption sets etc.). And this is exactly our goal in implementation theory and mechanism design trying to implement socially optimal decisions in economics. Therefore it is of relevance to look at social choice correspondences.

Definition 16 *A social choice correspondence, scc, is a mapping, $c : L(A)^M \curvearrowright A$, which assigns a subset $A' \in 2^A$ (the set of subsets in A) to any profile of individual preferences in $L(A)^M$.*

In welfare economics the most common used scc's are the Pareto-correspondence, and the individual rational correspondence, but here (in the course on Allocation Mechanisms) we need some more, e.g. (defined elsewhere in the examination requirements):

- the envy-free correspondence
- the fair-in-production correspondence, and
- the egalitarian-equivalent correspondence,

which all focus on (end state) equality in the population of agents. And of course we can combine different correspondences in different ways, and get e.g. the Pareto optimal and individual rational correspondence etc. We just have to ensure ourselves, every time we combine more correspondences or construct new ones, that there exists feasible states of the kind defined by the correspondence in the environment we focus on, e.g. an Arrow-Debreu economy of some specific sort.

In general, when we look at scc's there are a lot of possibility results connected to specific equilibrium concepts, and there are some impossibility results too. The possibility results, that are the most interesting results for mechanism design, and the impossibility results too, are part of game theory (see the next section) and implementation theory, see Corchón (1996).

3
Game Theory

Whenever a collective decision has to be made in a society, conflicts are immanent because different agents may have different opinions on matters of common concern and various means of influencing the common decision. In chapter 2 above we saw that also strategic behavior is an immanent part of collective decisions. So, if we want to design collective choice rules which implement socially optimal decisions, we must take these features of social choice procedures into account in the design of collective choice rules. In game theory it is possible to do so.

A social game consists of a game form defining the rules of the game, and a profile of characteristics (preferences) assigning to each agent e.g. a linear preorder of the alternatives in A, $(\succ_i)_{i \in M}$.

Definition 17 *A game form is an array, $G = [(S_i)_{i \in M}, A, g]$, where*

(i) $\forall i \in M : S_i$ is a non-empty set of legal (in the game) strategies.

(ii) A is a non-empty set of outcomes.

(iii) $g : (S_1 \times ... \times S_m) \curvearrowright A$ is a function (the outcome function).

- E.g. in "go fish" each player is given a strategy set, $S_i = (s_i^{1,1}, ..., s_i^{13,(m-1)})$, where $s_i^{1,1}$ is player No. $i (\neq 1)$ asking player No. 1: "all your aces" etc.; A is the set of alternative results of the game - a distribution of 13 tricks on m agents (each agent gets $g_i(S_1 \times ... \times S_m) \in [0;13]$ tricks, where $g = (g_i)_{i \in M}$ and $\sum_{i \in M} g_i(.) = 13$).

A game form lay down the rules to be followed by the parties in a conflict: the strategy sets comprising all legitimate acts (complete plans of actions) for the participants, and the outcome function assigning consequences to alternative acts of the agents. However, a game form in itself is insufficient to prescribe the actions of the agents. How an agent should or how he actually will act in a given collective decision situation depends not only on the rules of the game but also on the players' characteristics. As in social choice theory we assume here that the agents are characterized solely by their preferences. Thus we have:

Definition 18 *A game (in normal form) Γ is a game form, $G = \left[(S_i)_{i \in M}, \mathbb{R}^M, g\right]$, where $A = \mathbb{R}^M$, and a preference profile, $(u_i)_{i \in M}$, where for all $i \in M$ each $u_i : A \curvearrowright \mathbb{R}$, such that $\Gamma(G, (u_i)_{i \in M}) = \left[(S_i)_{i \in M}, \mathbb{R}^M, f\right]$ - or short: $[(S_i)_{i \in M}, f]$ - where the outcome function, $f(s_1, ..., s_m) = (u_1(g(s)), ..., u_m(g(s)))$, for $s = (s_1, ..., s_m) \in \Pi_{i \in M} S_i$.*

In the definition above we use the normal form description of a game. Another description of the same game is the extensive form representation, where the finer structure of the game is described (moves, information about the history of the game etc.). Thus we have

Definition 19 *An extensive representation of a game specifies for each player in the game, $(u_i)_{i \in M}$,*

(i) when each player has the move, what this player can do at each of his opportunities to move, and what information this player knows at each of his moves,

(ii) the set of outcomes, A, and

(iii) the outcome function.

Compared with the representation of games in normal form, we need to specify the strategy set of each agent in much more detail when the game is represented in extensive form: a strategy for each agent is here an array of moves, as it is known from the rules of the game in e.g. chess, bridge, "go fish" etc.

Any game (static or dynamic) can be represented in either normal or extensive form. A standard procedure representing every game in extensive form as a game in normal form exists, see Keiding (1987). But because you need more information to represent a game in extensive than in normal form, more games in extensive form are represented by the same game in normal form. Examples on how to change the representation of static games from normal to extensive form, and dynamic games from extensive to normal form, you can find in Gibbons (1992), chapter 2.4.

We consider the question whether we can select from the strategy set of each of the agents some "good" strategies (hopefully a single one), as more likely to be chosen by the agents than the rest. We only look at non-cooperative game theory. Also, we do not consider Bayesian equilibrium concepts. Instead we look at equilibrium i dominant strategies, Nash-equilibrium, trembling-hand perfect-, and subgame perfect Nash-equilibrium.

But what is an equilibrium? We use the following:

Definition 20 *Let* Γ *be a game. A strategy profile* $(s_1, ...s_m)$ *describing a complete plan of actions for every agent, is said to be an equilibrium, if and only if no agent ex-post regret his choice of strategy.*

- In comparative static environments a strategy profile is an equilibrium, if no agent has any positive reason to regret his choice of strategy when he receive his pay-off. In dynamic environments this definition need to be extended, because an equilibrium here is a (time dependent) path of states.

3.1 Dominant Strategy Equilibrium

It is not always easy to find such a subset of "good" equilibrium strategies. But in certain games there are some obvious candidates. One of these is a dominant strategy equilibrium.

Definition 21 *Let* Γ *be a game. We say that* s_i^0 *is a dominant strategy for agent* i *if for all* $s_i \in S_i$ *and all* $s_{)i(} \in \Pi_{)i(}S_{)i(}$, $f_i(s_i^0, s_{)i(}) \geqq f_i(s_i, s_{)i(})$. *Then* $s^0 = (s_1^0, ..., s_m^0)$ *is a dominant strategy equilibrium if* s_i^0 *is a dominant strategy for all* $i \in M$.

- No matter what strategy the other agents choose, $s_i^0 \in S_i$ will be a best strategy for agent i. And to calculate a dominant strategy (if it exists) the agent needs not to know the pay-off to other agents (their preferences or utility functions). In real life this means that the agents can choose a dominant strategy without knowing the other agents' choice of strategy, and since this will be relatively easy for an agent to do, a dominant strategy equilibrium is said to be a strong equilibrium concept.

Do games with a dominant strategy equilibrium exists? Yes, e.g. in Prisoners' Dilemma[1] (see Gibbons (1992), Lindeneg (2000), chapter 2) - a surprising feature in the Prisoners' Dilemma is, that the outcome of the game played in dominant strategies is a dominated alternative (both agents could be better off, if they choose another pair of strategies, but to do so they need to cooperate).

Do games in general have solutions in dominant strategies? Unfortunately not, e.g. the Battle of the Sexes (see Gibbons (1992), Lindeneg (2000), chapter 2). An easy way to examine whether a game has a dominant strategy equilibrium, is to check-up if every agent in the game has a dominant strategy. If so, then the game according to definition 21 has a dominant strategy equilibrium.

Examples of economic games with a dominant strategy equilibrium are:

[1] Originated in Tucker (1950).

- A market given by a price vector in an Arrow-Debreu model (see Lindeneg (2000), chapter 4).

- The Clarke-Groves mechanism allocating public goods in an environment with quasi-linear utility functions (see Corchón (1996), chapter 3.4.1).

But unfortunately in many economic games there are no dominant strategy equilibrium. Therefore we move on and look at some weaker equilibrium concepts: Nash-equilibrium and some refinements of that concept.

3.2 Nash Equilibrium

The most used equilibrium concept in economics in the last twenty years is Nash equilibrium (originally introduced by Nash (1950)). Agents have the possibility to choose either pure or mixed Nash-strategies. When looking at pure strategies we initially ignore the possibility that players might randomize over their pure strategies.

Definition 22 *Let* $\Gamma\left[(S_i)_{i\in M}, f\right]$ *be a game. A strategy set,* $s^* = (s_1^*, ..., s_m^*) \in (S_i)_{i\in M}$ *is a Nash equilibrium if for each agent,* $i \in M$, $f_i(s_i^*, s_{)i(}^*) \geqq f_i(s_i, s_{)i(}^*)$, *all* $s_i \in S_i$.

In a Nash equilibrium, each agent's strategy is a best response to the strategies actually chosen by the other agents. This implicate that an agent in *a simultaneous-move game* only ex-post (at the end of the game) knows if his own choice has been optimal for himself. And because all agents have to choose a strategy ex-ante they often can make a best choice only by chance, if they do not have complete and perfect information ex-ante.[2] Because of these strong informational requirements, Nash equilibrium is considered a weak equilibrium concept.

Do games in general have a Nash equilibrium? Fortunately the answer is "yes" under fairly broad circumstances. There ex-

[2] This is not necessary in a dynamic game if the equilibrium is stable, see e.g. Lindeneg (2000) chapter 6.

ist several existence results (see e.g. Mas-Colell *et al.* (1995) chapters 8 and 9), but here we look only at two of them.

Theorem 5 *Let* Γ *be a game where for all* $i \in M$:

(1) S_i *is a non-empty, convex, and compact subset of some Euclidian space* $\mathbb{R}^n$ *(*$n \in \mathbb{N}$ *(natural numbers) is the dimension of the strategy set).*

(2) $u_i(s_1, ..., s_m)$ *is continuous in* $(s_1, ..., s_m)$ *and quasiconcave in* s_i.

Then the game Γ *has a Nash equilibrium.*

For the proof we need a lemma:

Lemma 6 *If the sets* $S_1 \times ... \times S_m$ *are non empty,* S_i *is compact and convex, and* $u_i(.)$ *is continuous in* $(s_1, ...s_m)$ *and quasiconcave in* s_i*, then player* i'*s best-response correspondence* $b_i(.)$ *is non-empty, convex-valued and upper hemicontinuous.*[3]

Proof of the lemma: b_i is non empty because $b_i(., s_{)i(})$ is the set of maximizers on $u_i(., s_{)i(})$ on the set S_i. And b_i is convex because the set of maximizers on a quasiconcave function, $u_i(., s_{)i(})$, on a convex set, S_i, is convex. Finally, if any sequence $\left(s_i^\ell, s_{)i(}^\ell\right) \rightarrow \left(s_i, s_{)i(}\right)$ such that $s_i^\ell \in b_i(s_{)i(}^\ell)$ for all ℓ, we have $s_i \in b_i(s_{)i(})$, then $b_i(.)$ is upper hemicontinuous. To see this, note that for all ℓ, $u_i(s_i^{*\ell}, s_{)i(}^\ell \geqq u_i(s_i, s_{)i(})$ for all $s_i \in S_i$. And therefore by the continuity of $u_i(.)$ we have $u_i(s_i^*, s_{)i(}) \geqq u_i(s_i, s_{)i(})$. $\square$

And then we have

Proof of theorem 5: Define the correspondence $b : S \curvearrowright S$, from the non-empty, convex, and compact set, $S = S_1 \times ... \times S_m$, of strategy profiles to itself, by $b(s_1, ..., s_m) = b_1(s_{)1(}) \times ... \times b_m(s_{)m(})$. By using lemma 6 we have that all the conditions of the Kakutani fixed point theorem are satisfied, saying that a strategy profile exist, $s \in S$, such that $s \in b(s)$, where the

[3] A correspondence, $c : A \curvearrowright Y$, from a set $A \subset \mathbb{R}^N$ to a closed set, $Y \subset \mathbb{R}^K$, is upper hemicontinuous if it has a closed graph (given A and Y, c has a closed graph if for any two sequences, $x^\ell \rightarrow x \in A$, and $y^\ell \rightarrow y$, with $x^\ell \in A$, and $y^\ell \in c(x^\ell)$ for every ℓ, we have $y \in c(x)$), and the image of compact sets are bounded (for every compact set $B \subset A$ the set $c(B) = \{y \in Y : y \in c(x) \text{ for some } x \in B\}$ is bounded).

strategies constitute a Nash equilibrium by construction, $s_i \in b(., s_{)i(})$ for all $i \in M$. And then we are done with the theorem. □

Theorem 5 is a significant result, the requirements are satisfied in a wide range of economic applications. If applied to many examples in this paper, we should note that a finite strategy set cannot be convex. Therefore, in theorem 7 we look at Nash equilibria in mixed strategies, where agents assign lotteries on the pure strategies in their strategy set: it convexifies agents' strategy sets and yields well-behaved pay-off functions. We have:

Definition 23 *A mixed strategy profile, $\sigma = (\sigma_1, ..., \sigma_m)$ constitutes a Nash equilibrium of the game $\Gamma\left[(S_i)_{i\in M}, f\right]$, where for all $i \in M$: S_i is a convex set of strategies for agent i, and $f = (u_1, ..., u_m)$, if for every $i \in M : u_i(\sigma_i, \sigma_{)i(}) \geqq u_i(\sigma_i^{'}, \sigma_{)i(})$ for all $\sigma_i^{'} \in S_i$.*

Theorem 7 *Every game Γ in which the strategy sets, $S_1, ..., S_m$, have a finite number of elements has a mixed strategy Nash equilibrium.*

Proof: The game $\Gamma\left[(S_i)_{i\in M}, f\right]$ in definition 23 and theorem 7 satisfied all the assumptions in theorem 5 above. Hence theorem 7 is a direct corollary of theorem 5. □

The best argument for Nash equilibrium is that ex-post no agent will regret his own choice, see the discussion in Mas-Colell *et al.* (1995), pp. 248-49.

Examples of economic games with Nash-equilibria are:

- Schmeidler (1980)'s market mechanism in an Arrow-Debreu model (see Lindeneg (2000), chapter 5).

- The (stable) Cournot-Walras mechanism (Walker (1984)) in big Arrow-Debreu models (see Lindeneg (2000), chapter 6).

- The Cournot-Lindahl mechanism (Walker (1981)) in an Arrow-Debreu model with public goods (see Corchón (1996), chapter 5.2, and Lindeneg (2000), chapter 5).

Even if Nash equilibrium in the 1980's and (the begining of) the 1990's has been the most widely used solution concept in applications of game theory to economics, it turns out that this concept has some shortcomings when used in the design of allocation mechanisms:

a) In economic games there are sometimes too many Nash equilibria, and

b) The informational requirements - the agents need complete and perfect information - are very strong. And because many allocation mechanisms implementing socially optimal decisions in Nash equilibria are not stable, we need these requirements very often.

Therefore, in order to solve problems of type a) and b), we turn to look at some refinements of Nash-equilibrium which cuts down the number of equilibria and/or are stable in economic environments when we use allocation mechanisms.

3.3 Trembling-Hand Perfection

In this section we look at a perfection of Nash equilibrium which identifies Nash equilibria that are robust to the possibility that, with some small probability, the agents make mistakes and not choose a Nash equilibrium together. This equilibrium concept, see Selten (1975), is called (normal form) trembling-hand perfect Nash equilibrium, and can be used in voluntary contribution games, see e.g. Bagnoli and McKee (1991) - in Samfundsøkonomisk allokering 1998/1: Materialesamling.

For any normal form game, $\Gamma = [(S_i)_{i \in M}, f]$, we can define a perturbed game, $\Gamma_e = [(S_i^e)_{i \in M}, f]$, by choosing for each agent $i \in M$ and strategy $s_i \in S_i$ a number $e_i(s_i) \in [0;1]$, with $\sum_{i \in M} e_i(s_i) < 1$, and then define agent i's perturbed strategy set to be

$$S_i^e = \left\{ \sigma_i^P : \sigma_i(s_i) \geqq e_i(s_i) \text{ for all } s_i \in S_i \text{ and } \sum_{s_i \in S_i} \sigma_i(s_i) = 1 \right\}.$$

- A perturbed game is derived from the original game by requiring that each agent plays everyone of his strategies with at least some minimal positive probability, $e_i(s_i)$ - the unavoidable probability that strategy s_i gets played by mistake - for all $s_i \in S_i$.

Then we consider as predictions in game Γ only those Nash equilibria $\sigma = (\sigma_1, ..., \sigma_m)$ in the perturbed game Γ_e that are robust to the possibility that agents make mistakes. To consider σ as a robust equilibrium there must be at least some slight perturbations of Γ whose equilibria are close to σ. And we have:

Definition 24 *A Nash equilibrium σ of a game $\Gamma = [(S_i)_{i\in M}, f]$ is (normal form) trembling-hand perfect if there are some sequence of perturbed games $\{\Gamma_{e^k}\}_{k=1}^{\infty}$ that converges to Γ (i.e $\lim_{k\to\infty} e_i^k(s_i) = 0$ for all $i \in M$ and $s_i \in S_i$), for which there are some associated sequence of Nash equilibria $\left\{\sigma^k\right\}_{k=1}^{\infty}$ that converges to σ (such that $\lim_{k\to\infty} \sigma^k = \sigma$)*

Note we require only that some perturbed games exist that have equilibria arbitrarily close to σ. This is a mild test of robustness - a stronger test would be if we required σ to be robust to all perturbations close to the original game Γ.

Which games have trembling-hand perfect Nash equilibria? We have the following:

Theorem 8 *Every game, $\Gamma = [(S_i)_{i\in M}, f]$, with finite strategy sets, $S_1, ..., S_m$, has a trembling-hand perfect Nash equilibrium.*

Proof: See Selten (1975).

The set of trembling-hand perfect Nash equilibria, $(\sigma_i^P)_{i\in M}$, in a game with finite strategy sets is a subset of the set of undominated Nash equilibria in mixed strategies in this game ($e_i(s_i) = 0$ for any dominated strategy).

Above we have discussed trembling-hand perfection in normal form games. A slightly different form of trembling-hand perfection for dynamic games is discussed in Mas-Colell *et al.* (1995) chapter 9, Appendix B.

3.4 Subgame Perfect Nash Equilibrium

In section 3.1 - 3.3 we have studied simultaneous-move games. These can be used in comparative static models where time has no influence, e.g. in the environments of traditional Arrow-Debreu models. But in dynamic economic models these games will not be adequate to describe the behavior of the agents. Examples are models of repeated economic activity (most economic activities are repeated again and again), and extensive form games, see e.g. the compensation mechanism developed by Varian (1994), in Corchón (1996), chapter 6, Appendix I, and Lindeneg (2000) chapter 6.

In dynamic games we need other equilibrium concepts. Let us begin by illustrating that the Nash equilibrium concept may not give sensible predictions in dynamic games.

Look at dynamic games with *complete and perfect information*[4], see Gibbons (1992) section 2.1, and Mas-Colell *et al.* (1995) chapter 9.B. Consider a game played by two firms, a monopolist (M), and an entrant (E). E has two (pure) strategies, (a) entering the market, or (b) stay out of the market. If E chooses (a) then M has two (pure) strategies, (α) to fight E, or (β) to share the market with E, and if E chooses (b) nothing happens (status quo). The pay-off matrix is:

Table 1:

$E \diagdown M$	α	β
a	$-3, -1$	$2, 1$
b	$0, 2$	$0, 2$

The normal form representation of this game has two pure strategy Nash equilibria: $(s_E, s_M) = (a, \beta)$ or (b, α). If E enters the market the best strategy for M is to share the market. But the only chance that E stays out of the market is if M declares to fight, if E enters the market, and in that case status quo remains. Yet the stragegy profile, $(s_E, s_M) = (b, \alpha)$, is not cred-

[4] In a game with complete information every agent knows the game form. And in a game with perfect information every agent knows exactly what strategy every other agent in the game will choose.

ible, because if E has complete and perfect information it can foresee that if it enters the market then M will choose to share the market, and in this case, $(s_E, s_M) = (a, \beta)$, E will increase its outcome from 0 to 1. Therefore it will be a dominant strategy for E to choose strategy a and enter the market.

This example illustrates the problem in using the Nash equilibrium concept in dynamic games: it permits M to make an empty threat that E nevertheless takes seriously when choosing it's strategy. To rule out strategies as (b, α) we can insist that agents' equilibrium strategies satisfied the principle of *sequential rationality*: an agent's strategy must specify optimal actions at every decision node in the extensive form game tree. That is, if an agent is at some decision node in the game tree, his strategy should prescribe credible actions from that node on given the other agents' strategies. In the example above M's strategy "fight if E chooses to enter the market" does not: if E enters the market the only optimal strategy for M is to share the market with E.

There exist a procedure to ensure that a sequential rational Nash equilibrium is chosen. This procedure, backward induction, first solve for optimal behavior at the last decision node of the game tree, then in the second to the last etc. If we first determine optimal behavior for M at the post entry stage of the game in the example above, the choice will be $s_M = (\beta)$ "share the market". And then we solve at the first node, knowing that M in the next step will choose (β), E chooses (a) "to enter the market", see figure 1.

Figure 1:

E

out　　　　in

(E,M)=(0,2)　　　　M

fight　　　　share

(E,M)=(-3,-1)　　(E,M)=(2,1)

The idea of backward induction works in all finite games of perfect information. We have:

Theorem 9 *(Zermelo's theorem). Every finite game of complete and perfect information has a pure strategy Nash equilibrium that can be derived through backward induction. Moreover, if no agent has the same payoffs at any two terminal nodes, then there is a unique Nash equilibrium that can be derived in this manner.*

Proof: See Mas-Colell *et al.* (1995), pp. 272-73.

This solution concept, sequential rational Nash equilibrium, has been generalized to games of *complete but imperfect information* by Selten (1965), in the concept of subgame perfect Nash equilibrium. To define this equilibrium concept we must know what a subgame is.

Definition 25 *A subgame of an extensive form game* Γ_{Ext} *is a subset of the game having the following properties:*

a) It begins with an information set containing a single decision node plus all the decision nodes that are successors of this node, and contains only these nodes.

b) If a decision node is in the subgame, then every decision node in the same information set as this node is also in the subgame (no broken information sets).

E.g. the game in figure 1 has two subgames: the game as a whole and the part of the game tree that begins with and follows M's decision node (note that in a finite game of perfect information, every decision node initiates a subgame).

Below we say that a strategy profile σ induces a Nash equilibrium in a particular subgame of an extensive form game, if the moves specified in σ for information sets within the subgame constitute a Nash equilibrium for the subgame and nothing but this subgame. Then we have:

Definition 26 *A strategy profile* σ *in an extensive form game,* $\Gamma_{Ext.}$, *is a subgame perfect Nash equilibrium if it induces a Nash equilibrium in every subgame of* $\Gamma_{Ext.}$.

Note that the set of subgame perfect Nash equilibria (SPNE) is a subset of the set of Nash equilibria (NE), $SPNE \subseteq NE$.

And that the set of $SPNE$ equals the set of NE derived through backward induction in games of perfect information.

From Zermelo's theorem we get directly:

Corollary 10 *Every finite extensive form game of complete and perfect information has a pure strategy subgame perfect Nash equilibrium. Moreover, if no agent has the same payoffs at any two terminal nodes, then there is a unique $SPNE$.*

And in the case of complete but imperfect information, that

Theorem 11 *Any finite extensive form game of complete information has a subgame perfect Nash equilibrium in pure or mixed strategies if a Nash equilibrium is present.*

Proof: The argument is by construction. If we generalize the backward induction procedure as follows:

1. Start at the terminal decision nodes, and identify the Nash equilibria for each of the final subgames.

2. Select one Nash equilibrium in each of these final subgames, and derive the reduced extensive form game in which these final subgames are replaced by their payoffs when agents choose these Nash strategies.

3. Repeat step 1 and 2 for the reduced game. Continue the procedure until every move in the original extensive form game is determined. Then the sample of moves at the various information sets of the game constitute a profile of $SPNE$ strategies.

4. If multiple equilibria exist then the full set of $SPNEs$ is identified by repeating the procedure for each possible equilibrium that could occur for the subgames in question.

Because a Nash equilibrium exists in all (static or dynamic) finite normal form games of complete information that fulfil (1) and (2) of theorem 5, and any finite dynamic game of complete information has a finite number of subgames that satisfies (1) and (2) of theorem 5, then the procedure above can identify at least one subgame perfect Nash equilibrium in a finite extensive form game of complete information. □

Note that the proof is constructive: if we use the procedure 1-4 it is possible to determine the set of $SPNEs$. And that $SPNE$

eliminates Nash equilibria that rely on non-creditable threats or promises. This is used in mechanisms implementing socially optimal decisions in economic environments with complete but imperfect information in subgame perfect Nash equilibrium, see e.g. Lindeneg (2000), chapter 6.

3.5 Other Equilibrium Concepts

Besides the above mentioned equilibrium concepts the examination requirements in Allocation Mechanisms include af few more non-cooperative, non-Bayesian equilibrium concepts. These are other refinements of Nash equilibrium, namely *undominated Nash equilibrium* (see Corchón (1996) section 6.3) which choose an allocation among the non-dominated Nash equilibria, and *virtual implementation* choosing a Nash equilibrium ε-close to the correspondence we want to implement (see Corchón (1996) section 6.4).

And it is also possible to find a single definition of a cooperative equilibrium concept, strong (or Aumann) equilibrium in Corchón (1996) section 2.6, definition 5. But forget all about this latter concept - we do not use cooperative equilibrium concepts in this cource.

4
Allocation Mechanisms

In this last section we make a survey of some allocation mechanisms central in the course on Allocation Mechanisms at the University of Copenhagen, see table 2.

Table 2:

Mechanism name:	$\mathbf{P} : \mathbf{E} \curvearrowright \mathbf{A}$	**Equilibrium**	**Stability**
a) $Clarke - Groves$	$\max \sum_{i \in M} u_i$	dominant	$-$
b) $Schmeidler - market$	$PO \cap IR$	$Nash-$	No
c) $Cournot - Lindahl$	$PO \cap IR$	$Nash-$	No
d) $Cournot - Walras$	$PO \cap IR$	$Nash-$	$best - replay$
e) $Compensation$	$PO \cap IR$	Subgame-perfect	$best - replay$
f) $Combination$	$PO \cap IR$	Subgame-perfect	$best - replay$
g) $Voluntary - contrib.$	$PO \cap IR$	Trembling-hand	In practice

a) The Clarke-Groves mechanism, see Clarke (1971) and Groves (1973), implements the utilitarian decision rule in dominant strategies when it is possible to make interpersonal comparisons of different agents' utilities (this is needed whenever the utilitarian decision rule is used) and quasilinear preferences. Compared with the Arrow-conditions in theorem 1 above this is a restriction on the domain for the preferences of the agents (which is in conflict with the rather implicit assumption in section 2.1 of unrestricted domain in Arrows impossible theorem).

b) The market mechanism developed by Schmeidler (1980) implements the Walrasian correspondence in Arrow-Debreu environments in Nash equilibrium. The Walrasian correspondence is Pareto optimal and individual rational in Arrow-Debreu economies without market failures. However, there is an additional condition (connected with the conditions in the Arrow-Debreu model without market failures) for the mechanism to work like this, namely that $m \geqq \ell + 1$ (the number of agents must be larger than the number of commodities). And, furthermore, the mechanism is unstable out of equilibrium.

c) The Cournot-Lindahl mechanism developed by Walker (1981) implements the Lindahl-correspondence in Arrow-Debreu environments with public goods in Nash equilibrium when the number of agents in the economy, $|M| \geqq 3$, and if it is combined with a market mechanism implementing the Walrasian correspondence in all markets without marketfailures. For practical applications it is important to note that the Nash equilibria in this mechanism are not stable.

d) The Cournot-Walras mechanism in Walker (1984) implements the Walrasian correspondence in big (the number of agents $= \infty$) Arrow-Debreu models without market failures in Nash equilibrium. The Nash equilibria are stable if and only if the agents use Cournot best-replay behavior when an economy is out of equilibrium.

e) The Compensation mechanism developed by Varian (1994) implements the Pareto optimal and individual rational correspondence in subgame perfect Nash equilibrium in (dynamic) Arrow-Debreu environments with externalities. The mechanism is stable if the agents use best-replay behavior out of equilibrium.

f) The Combination mechanism, see Lindeneg (2000) section 6*5, combines the compensation and the Cournot-Walras mechanisms in one mechanism that implements the Pareto optimal and individual rational correspondence in (dynamic) Arrow-Debreu environments with market failures (exclusive some types of asymmetric information). The mechanism is stable if the agents in the economy use best-replay behavior out of equilibrium.

g) The Voluntary Contribution mechanism, see Bagnoli and McKee (1991), implements the Pareto optimal and individual rationel correspondence in trembling-hand perfect Nash equilibrium in a (partial analysis) market of one specific good in a given number. Experiments with students show that equilibria when using this mechanism are stable. In equilibrium the mechanism is balanced (budget neutral). Because the contributions from the agents are voluntary this mechanism is individual rational. This means that the Voluntary Contribution mechanism can be used to implement the Pareto optimal and individual correspondence in an Arrow-Debreu economy containing public goods delivered in discrete amounts.

Besides these mechanisms the examination requirements include a few other types of mechanisms: auctions (see Rasmussen in Materialesamling til Samfundsøkonomisk allokering (1998)), the cost-share mechanism (see Corchón (1996) section 5.3), and the 'divine and permute' mechanism (see Corchón (1996) section 5.4).

5
References

Arrow, K.J., Social choice and individual values, Wiley 1951 (2nd ed. 1963)

Bagnoli, M. and M. McKee (1991), Voluntary contribution games: efficient private and public provision of public goods. Economic Enquiry **29**: 351-366

Borda, J.C. de, Memoires sur les élections au scrutin. Histoire de l'Akademie Royale des Sciences, Paris 1781

Condorcet, Marquis de, Essai sur l'appication de l'analyse á la probabilité des décisions rendues á la pluralité de voix. Paris 1785

Clarke, E. (1971), Multipart pricing of public goods. Public Choice **11**: 19-33

Corchón, L.C., The theory of implementation of socially optimal decisions in economics. MacMillan Press 1996

Gibbard, A. (1973), Manipulation of voting schemes: A general result. Econometrica **41**: 587-601

Gibbons, R., A primer in game theory. Harvester-Wheatsheaf 1992

Groves, T. (1973), Incentives in teams. Econometrica **41**: 617-

631

Keiding, H. (1984), Planning theory, part III. Lecture notes, Institute of Economics, University of Copenhagen (stencil)

Keiding, H., Spilteori. Akademisk Forlag 1987

Lindeneg, K., Frihed, lighed & effektivitet? Akademisk Forlag 2000

May, K.O. (1952), A set of independent necessary and sufficient conditions for simple majority decisions, Econometrica **20**: 680-684

Mas-Colell, A., M.D. Whinston and J.R. Green, Microeconomic theory. Oxford University Press 1995

Moulin, H., The strategy of social choice. North-Holland 1983

Muller, E. and M.A. Satterthwaite (1977), The equivalence of strong possitive association and strategy proofness. Journal of Economic Theory **14**: 412-418

Nash, J. (1950), Equilibrium points in n-person games. Proceedings of the National Academy of Sciences **36**: 48-49

von Neumann, J. and O. Morgenstern, Theory of games and economic behavior. Princeton University Press 1944

Satterthwaite, M.A., The existence of strategy-proof voting procedures, a topic in social choice theory. Ph.D. dissertation, University of Visconsin, Madison1975

Schmeidler, D. (1980), Walrasian analysis via strategic outcome functions. Econometrica **48**: 1585-1593

Selten, R. (1965), Spieltheoretische behandlung eines oligopolmodells mit nachfragetragheit. Zeitschrift für die gesamte Staatswissenschaft **121**: 301-324

Selten, R. (1975), Re-examination of the perfectness concept for equilibrium points in extensive games. International Journal of Game Theory **4**: 25-55

Tucker, A.W. (1950), A Two-Person Dilemma, working paper, Standford University

Varian, H.R. (1994), A solution to the problems of externalities when agents are well-informed. American Economic Review

84: 1278-1293

Walker, M. (1981), A simple incentive compatible scheme for attaining Lindahl allocations. Econometrica **49**: 65-71

Walker, M. (1984), A simple auctioneerless mechanism with Walrasian properties. Journal of Economic Theory **32**: 111-127